**DRAFT Guidelines for inclusion of profiling float data into the Argo Data Stream**

**Executive Summary**

The increasing size and complexity of the Argo data stream has placed enormous pressure on the Argo data management system. As a result, it is necessary to develop guidelines for floats and sensors that can be submitted to the Argo Data Stream. The Argo Program has been supported by the International Oceanographic Commission to provide a high quality, global data set for tracking variations on seasonal to inter-annual time and 1000 km space scales while respecting its member state’s territorial rights. While national DACs (Data Assembly Centers) must use their judgement to accept a float into the Argo data stream, the AST and ADMT have provided this document to assist them in making their determination.

To maximize the use of the Argo Array and provide the necessary high-quality data needed for climate research, a float included in Argo must:

* follow the Argo governance rules for pre-deployment notification and timely data delivery of both real-time and delayed-mode quality controlled data.
* have a clear plan for long term data stewardship through a national Argo Data Assembly Centre.
* should target the Core Argo profiling depth and cycle time, 2000 m and 10 days, respectively. However, data that contributes to the estimation of the state of the ocean on the scales of the Core Argo Program are also desirable.

The Argo Program encourages development and improvement of sensors that measure climate related parameters. The path for full-acceptance of a new sensor into the Argo Data Stream should include:

* Experimental deployments – either on an ‘Argo float’ with the data from these sensors only included in auxiliary files, or outside of Argo on a research float.
* An AST approved pilot study, with the data included in the profile files, but grey listed and appropriately flagged to prevent an impact on users
* AST acceptance that the data from the new sensor meet Argo accuracies, with the data included in standard form.

**Development of New Sensors**

As the use of profiling floats increases and their data become more complex, the Argo Program must balance the benefits of improved coverage with the increased complexity and difficulty of documenting, managing and distributing Argo data. After official notification prior to deployment, an ‘Argo float’ produces freely available, real-time quality controlled data within 24 hours. The data is further quality controlled and assembled into a long-term archive for use in critical scientific analysis. This requires standards for sensor data quality and accuracy and for responsibilities for notification and quality control. Novel sensors further complicate the management of Argo floats.

To help national programs determine whether floats with different sampling regimes or sensors should be included in the Argo program, the Argo Steering Team and Argo Data Management Team have developed a list of requirements for a float to be an ‘Argo Float.’

The Argo Program encourages the appropriate development of new sensors for presently accepted parameters, and of new sensors for new parameters. Sometimes the most effective way to do this is by adding experimental sensors on floats that also contribute to Argo. Since ALL measurements from an Argo float must be made publicly available, including experimental ones, it is not practical to incorporate all new sensor data into the Argo Netcdf formatted data. In order to manage the progress of sensors from experimental to fully approved, Argo considers new sensors and parameters to be in one of three stages:

**Approved**: The performance and accuracy of an approved sensor are fully characterized. There is a well-defined data quality control path for the data. Approved sensors appear in the Argo reference tables for sensor description and measured parameters.

**Pilot**: A sensor has been developed, for either an existing parameter, or to measure a new parameter, and has been shown to have the potential to meet Argo requirements for accuracy and quality control procedures. This sensor is also likely to be mounted on a significant fraction of the Argo fleet. A Pilot Study using these sensors should be proposed to the AST. This study should be large enough to qualify the new sensor over different oceanographic conditions. The necessary description of the sensor and the measured parameters, meta and technical data should be proposed to the ADMT. If the Pilot Study is approved, the data from these sensors will be stored in the main Argo NetCDF files, but they will be grey listed, with a quality flag of 3.

**Experimental**: An experimental sensor has been mounted on an Argo float with an accepted CTD sensor, but its performance and accuracy have not been determined. If mounted on an Argo float, the data from the accepted sensors are included at the GDACs in Argo NetCDF format. Experimental data are placed in the ‘aux’ directory at the GDAC. Experimental sensors can also be tested outside of Argo.

**Presently Acceptable data for the Argo Data Stream**

The AIC must be notified, including a full list of sensors, before any Argo float is deployed. At a minimum, Argo floats should measure temperature, salinity and pressure using ‘accepted’ sensors that can meet Argo accuracy requirements. Target cycle time, maximum profile depth and park depth are 10 days, 2000 dbar, and 1000 dbar, respectively. Accepted Argo sensors are identified in the Argo Reference Tables (http://tinyurl.com/nwpqvp2 ) and measure accepted Argo parameters as identified in Reference Table 3 of the Argo User’s Manual (http://www.argodatamgt.org/content/download/27444187206/file/argo-parameters-list-core-and-b.xlsx /). These ‘Argo parameters’ must meet general requirements related to data quality and management. At present (March 2018) The only *approved* CTD is the SBE41 operating to 2000 dbar. The SBE61, SBE41 deployed deeper than 2000 dbar, and RBR CTDs are pilot sensors.

Data from sensors with pilot status should be included in Argo Netcdf files, but it should also be added to the Grey list, have a real-time QC flag of 3 and not be submitted to GTS via BUFR. The QC flag can be improved in delayed-mode, if the Argo data system agrees that the data are good.

To satisfy the Argo requirement that all measurements are transparent and available to the public and to minimize the burden on the Argo data Management Team, Argo has formed an auxiliary directory at the Argo GDACs that will distribute, but not curate this data. Documentation describing the contents and means to access the data in these directories is mandatory and the responsibility of the float provider. If over time the experimental sensors are shown to meet Argo's requirements or form part of a new global mission, a case for approval needs to be made to the AST and ADMT to include these into the Argo data stream.

If a float carries an Experimental CTD, then it can only be part of Argo if it also carries an Accepted CTD. Data from the Accepted CTD will be in the Argo Netcdf files; Data from the Experimental CTD will be in the ‘aux’ directory. If a float carries a Pilot and Approved CTD, the data from the Approved sensor will populate the PRES, TEMP, PSAL parameters.

The. To facilitate this.

**Requirements for inclusion of Argo floats into the Argo data stream with approved sensors and approved parameters:**

|  |  |  |
| --- | --- | --- |
| **Requirement** | **Comment** | **Who in Argo can oversee this requirement?** |
| **General requirements** |
| The float and its data are consistent with Argo governance (IOC XX-6; IOC-EC\_XLI.4) | It is the PIs responsibility to ensure that floats are managed according to international governance of Argo’s activities. This includes notifying, through a national contact point via the AIC, float deployments and when floats drift into EEZs of specific coastal states that have asked for notification. [refer to table?] | AIC will monitor consistency with Argo governance. If the float does not conform, AIC will notify national points of contact, who will take whatever action is necessary. |
| The float has a national contact point for its entire lifetime, and for the data processing after the float is dead | The AIC requires a contact point for all floats identified as ‘Argo’ by their WMO number. Where individuals’ responsibilities change within national programs and research groups, contact points must be maintained. PIs must consider the long-term ownership of and responsibility for their floats. It is possible that ‘ownership’ of a float could be passed to a contact in a national program, but this should be planned before the float becomes part of Argo. | Points of contact in the national program of the PI deploying the float. Where AIC is uncertain of a contact point, AIC will refer to points of contact in the relevant national program. |
| The float PI will arrange safe recovery and disposal of the float if beached | Once a float is part of Argo, the PI or an identified delegate must remain responsible for the float throughout its lifetime. The float cannot stop being an Argo float just because it has drifted out of a PI’s area of interest, or fulfilled the PI’s initial purpose.  | AIC, supported by national points of contact. |
| **Data Quality and Management** |
| The data have an established end-to-end pathway | Before a PI proposes to their national program that a float will become an Argo float, there must be an agreed pathway for: * data telemetry to a real time (RT) DAC
* preparation of real time files for the GTS and submission to a GDAC through an existing DAC that include all telemetered parameters
* Real Time Quality Control (RTQC) procedures for all parameters
* a pathway to agreed Delayed Mode Quality Control (DMQC) procedures for all parameters
* a DMQC group identified and funded for all approved parameters
* a commitment by a DMQC group for long-term curation of the data, with the ability to respond to evolving requirements from ADMT.

This should be in place when a float is notified to the AIC and added to the Argo data system, which is the moment at which a float becomes an Argo float. | The person making the notification to AIC. If this is an established Argo point of contact, it is the notifying person’s responsibility to be sure everything is in place. If a float is notified to AIC from a new source, AIC should check that the PI is ready to fulfill their obligations and **is linked to a national Argo program**. |
| The data meet Argo targets for accuracy and vertical sampling characteristics | New sensors that measure parameters already established in the Argo system must have been demonstrated to meet Argo targets for accuracy. This can be established through peer-reviewed papers and or freely available pilot data sets. Vertical sampling must be deep enough and often enough for delayed-mode quality control to be performed. Where no published statements of requirement exist, AST or ADMT will evaluate proposals or develop statements of requirement. | AST advised by ADMT. |
| The PI has established a data management arrangement with a recognized Argo DAC and delayed-mode group, including the supply of all technical and metadata | The PI is advised to negotiate the pathway for RT and DM activities with one of the recognized Argo DACs before stating in funding applications that floats will be a contribution to Argo. This includes timely delivery of appropriate technical and metadata. Argo DACs may require funding for extra staff time to handle extra floats, sensors and parameters. | RT DACs will ensure the necessary arrangements are in place before starting to accept data from new floats. This should be agreed between RT DACs and float PIs before the float is deployed or notified. |
| Addition of new sensors and parameters on floats in the Argo data system requires endorsement from the AST & ADMT | Argo requires all data from all sensors on an Argo float to be made available at the Argo GDACs. The list of documented [Argo <PARAM> names](http://www.argodatamgt.org/content/download/27444/187206/file/argo-parameters-list-core-and-b.xlsx) (Reference Table 3 in the [Argo User's Manual](http://dx.doi.org/10.13155/29825)) cannot grow indiscriminately without undermining the Argo data system. AST and ADMT should only endorse the addition of new sensors and new parameters when there is a reasonable future prospect of those sensors being deployed in sufficient numbers and with appropriate global distribution. | AST advised by ADMT. |

In order to accept data for distribution by the Argo GDACs that originate beyond the core T/S and BGC data from floats deployed by national programs, a DAC must ascertain that:

|  |  |  |
| --- | --- | --- |
| There are sufficient resources to manage and archive these data in perpetuity according to evolving ADMT requirements | As Argo has evolved, the ADMT has agreed to a number of changes to the Argo data system. The most recent major change was the introduction of v3 format NetCDF files, with substantial impact on the contents and storage of profile, trajectory, technical and metadata. DACs have had to bear the burden of updating or converting legacy files in the Argo archive. Lesser examples include reprocessing data to review [TNPD floats](http://www.argo.ucsd.edu/Data_FAQ.html#TNPD) and requests for more complete metadata on sensors. The burden is disproportionately large for floats that lie outside the core T/S/P and, if applicable, the 6 core BGC variables. DACs and PIs need to consider the long-term responsibility for float data before a float can be accepted into the Argo system. | DACs will ensure the necessary arrangements are in place before starting to accept data from new floats. This should be agreed between DACs and float PIs before the float is deployed or notified. |
| A delayed-mode group has agreed to perform calibration and quality control in perpetuity | From time to time ADMT has asked DACs to revisit DMQC in the light of new insight into sensor behavior. Floats must have a plan for long-term responsibility, so that floats don’t become ‘orphans’. | DACs should ensure they know who is responsible for short- and long-term DMQC for all floats that fall in their directory at GDAC. This should be agreed between DACs and float PIs before the float is deployed or notified. |
| Sufficient meta data are provided by the PI. | PIs must undertake to provide to the responsible DAC whatever metadata are defined from time to time as mandatory by ADMT. | DACs should notify national points of contact of any floats that don’t conform to ADMT’s metadata requirements. |

**Use of the Auxiliary directory**

|  |  |  |
| --- | --- | --- |
| **Step** | **Action** | **Comment** |
| 1 | Agreed Argo parameters ([Reference Table 3](http://www.argodatamgt.org/content/download/27444/187206/file/argo-parameters-list-core-and-b.xlsx) of the Argo User's Manual) from approved Argo sensors (found in the [Argo Reference Tables](http://tinyurl.com/nwpqvp2)) should be handled through the DACs and GDACs in the usual way. The only approved Argo CTD sensor is the SBE41 to 2000db. SBE61, SBE41 below 2000db and RBR CTD are being piloted. | See the above tables for more details. |
| 2  | Agreed Argo parameters ([Reference Table 3](http://www.argodatamgt.org/content/download/27444/187206/file/argo-parameters-list-core-and-b.xlsx) of the Argo User's Manual) from pilot Argo sensors (found in the [Argo Reference Tables](http://tinyurl.com/nwpqvp2)) should be flagged as QC=3 until their accuracy is established and appropriate errors can be assigned.  | PIs should use the Reference Table 3 for this parameter . The Reference Tables will distinguish between pilot and approved sensors and parameters. |
| 3 | 3.1) Auxiliary (not in Argo Ref Table 3) parameters must be completely decoded by the PI. 3.2) The PI (or responsible DAC) must generate ASCII or NetCDF files that are easily read. PIs may be required to respond to enquiries for help in accessing data. 3.3) There must be a readme file that provides sufficient information on the contents of the PI files:  - what the sensor is  - what has been reported (parameters and units)  - how the data are organized | PIs are encouraged to use recognized [IODE names](http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P02) for new parameters wherever possible, if this can be done without ambiguity. Novel measurements may not have a suitable IODE name. Cross-checking with other major observing programs such as OCEANSites and GO-SHIP for already established parameter names is required.The SPECIAL\_FEATURES(1024) variable in the DAC meta.nc file must be filled by the DAC, to indicate there are AUX data files.Note: there is no SPECIAL\_FEATURES variable in single-cycle profile files, so there is no scope for a 1-1 pointer between single cycle profile files and auxiliary data files.  |
| 4 | PIs will work with a DAC to get the auxiliary files uploaded to the GDACs | Data served by GDACs in a 'parallel' aux directoryGDACs will provide inventory of filesSyntax of parameter names is chosen by PIs and could be updated at any time by PI. |

**Auxiliary file naming convention**

File names should use the following convention at GDAC:

***[WMO]***\_[cycle]\_[PIfield]\_***aux***.[nc|csv|txt]

***[WMO]***\_[PIfield]\_***aux***.[nc|csv|txt]

Bold italic part is mandatory and must match core file convention for WMO [and cycle].

**Auxiliary directory files and paths**

Example of files that would be available on GDACs for an Argo float with an acoustic and EM shear sensor.

**Core Argo files would appear on GDACs with the following pathways:**

Float files:

/pub/dac/bodc/1900083/1900083\_meta.nc - including SPECIAL\_FEATURES variable filled by DAC indicating there are auxiliary data files

/pub/dac/bodc/1900083/1900083\_tech.nc

/pub/dac/bodc/1900083/1900083\_prof.nc

/pub/dac/bodc/1900083/1900083\_Rtraj.nc

Core Argo profile for cycle one:

/pub/dac/bodc/1900083/profiles/D1900083\_001.nc

**Additional files for the auxiliary sensors would appear in the 'dacaux' folder on GDACs with the following pathways:**

Profile files:

/pub/dacaux/bodc/1900083/profiles/1900083\_001\_acoustic\_aux.csv

/pub/dacaux/bodc/1900083/profiles/1900083\_001\_emshear\_aux.csv

Examples of actual AUX files on the GDACS:

Float files:

/pub/dacaux/bodc/3901891/3901891\_tech\_aux.nc

/pub/dacaux/coriolis/5903126/5903126\_tech\_aux.nc

**SPECIAL\_FEATURES string**

The SPECIAL\_FEATURES string must include the substring ‘AUX’ for machine parsing. Examples:

‘This float has auxiliary sensors for acoustic measurement of rainfall’

‘This float has auxiliary sensors for optical measurement of oxygen’

 ‘This float has auxiliary sensors for electromagnetic measurement of vertical velocity shear’

‘This float has auxiliary sensors for turbulence measurements’

The AIC presently tracks the sensors that provide measurements in the PARAMETER list in the meta.nc file.

The AIC can also make the SPECIAL\_FEATURES string available.

The SPECIAL\_FEATURES string should be suitable to be included as part of notification of floats about to drift into EEZs. It must be clear and completely describe all auxiliary measurements.

**Acronyms**

AIC: Argo Information Centre – based at JCOMMOPS in Brest, the AIC independently monitors the location and status of all Argo floats for the member nations of the International Oceans Commission and the World Meteorological Organisation.

DAC: Data Assembly Centre – the national team responsible for encoding and exchanging Argo data with the Argo Global Data Asssemby Centres, and for insuring data broadcast on the GTS

GDAC: Global Data Assembly Centre – Argo has two, one based at IFREMER in France and one based at USGODAE in the USA. These are mirrored hourly

PI: Principle Investigator

IOC: International Oceans Commission, a subsidiary body of UNESCO.

RT: Real-time – a descriptor of the fast-delivered version of Argo data, typically within 24 hours or less.

DM: Delayed-mode – a descriptor of the highly quality controlled slow-delivered version of Argo data, typically within 1-2 years.

BGC: Bio-geo-chemical - referring to sensors/floats measuring chemistry and bio-optical parameters

QC: quality control – flagging and adjusting data from drifted senso